The crystal structure of the flavin containing enzyme dihydroorotate dehydrogenase A from *Lactococcus lactis*

Paul Rowland[1], Finn S Nielsen[2], Kaj Frank Jensen[2] and Sine Larsen[1]*

Background: Dihydroorotate dehydrogenase (DHOD) is a flavin mononucleotide containing enzyme, which catalyzes the oxidation of (*S*)-dihydroorotate to orotate, the fourth step in the *de novo* biosynthesis of pyrimidine nucleotides. *Lactococcus lactis* contains two genes encoding different functional DHODs whose sequences are only 30% identical. One of these enzymes, DHODA, is a highly efficient dimer, while the other, DHODB, shows optimal activity only in the presence of an iron-sulphur cluster containing protein with which it forms a complex tetramer. Sequence alignments have identified three different families among the DHODs: the two *L. lactis* enzymes belong to two of the families, whereas the enzyme from *E. coli* is a representative of the third. As no three-dimensional structures of DHODs are currently available, we set out to determine the crystal structure of DHODA from *L. lactis*. The differences between the two *L. lactis* enzymes make them particularly interesting for studying flavoprotein redox reactions and for identifying the differences between the enzyme families.

Results: The crystal structure of DHODA has been determined to 2.0 Å resolution. The enzyme is a dimer of two crystallographically independent molecules related by a non-crystallographic twofold axis. The protein folds into an α/β barrel with the flavin molecule sitting between the top of the barrel and a subdomain formed by several barrel inserts. Above the flavin isoalloxazine ring there is a small water filled cavity, completely buried beneath the protein surface and surrounded by many conserved residues. This cavity is proposed as the substrate-binding site.

Conclusions: The crystal structure has allowed the function of many of the conserved residues in DHODs to be identified: many of these are associated with binding the flavin group. Important differences were identified in some of the active-site residues which vary across the distinct DHOD families, implying significant mechanistic differences. The substrate cavity, although buried, is located beneath a highly conserved loop which is much less ordered than the rest of the protein and may be important in giving access to the cavity. The location of the conserved residues surrounding this cavity suggests the potential orientation of the substrate.

Addresses: [1]Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark and [2]Center for Enzyme Research, Institute of Molecular Biology, University of Copenhagen, Sølvgade 83H, DK-1307 Copenhagen K, Denmark.

*Corresponding author.
E-mail: sine@xray.ki.ku.dk

Key words: dihydroorotate dehydrogenase, flavoprotein, *Lactococcus lactis*, pyrimidine nucleotide biosynthesis, X-ray crystallography

Received: 18 September 1996
Revisions requested: 16 October 1996
Revisions received: 13 November 1996
Accepted: 4 December 1996

Electronic identifier: 0969-2126-005-00239

Structure 15 February 1997, 5:239–252

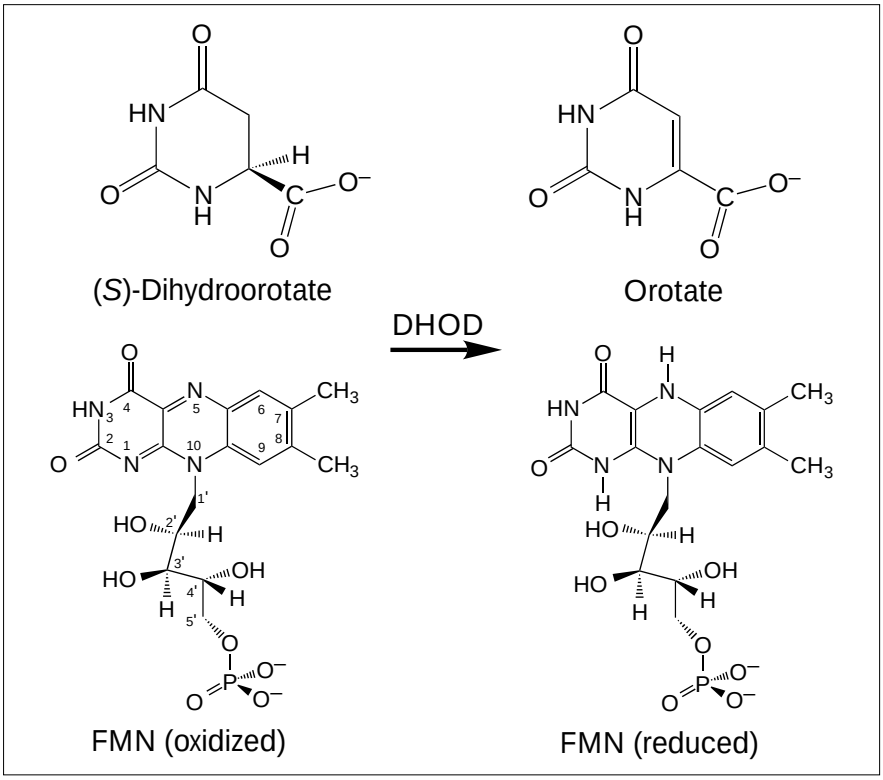
Introduction

Dihydroorotate dehydrogenase (DHOD; EC 1.3.3.1) is a flavin mononucleotide (FMN) containing oxidoreductase. It catalyzes the conversion of (*S*)-dihydroorotate to orotate, the fourth step and only redox reaction in the *de novo* biosynthesis of pyrimidine nucleotides. Oxidation of the enzyme substrate is accompanied by the corresponding reduction of the flavin group (Fig. 1). The reduced flavin can subsequently become reoxidized by various electron acceptors including dichlorophenolindophenol, potassium hexacyanoferrate (III) and molecular oxygen depending upon the enzyme source.

Genes encoding the enzyme have now been cloned and sequenced from a wide variety of species, revealing a diverse group of proteins. DHOD from *Escherichia coli*, which is known to be membrane bound [1], shows high sequence similarity to all DHODs of mitochondrial origin (greater than 40% identity), but very little similarity (less than 20% identity) with the corresponding enzymes from Gram-positive bacteria, such as the milk fermenting bacterium *Lactococcus lactis* [2] or the cytosolic enzyme from *Saccharomyces cerevisiae* [3].

It has recently been discovered that *L. lactis* contains two genes (*pyrDa* and *pyrDb*) encoding functional DHODs [2]. Both of these enzymes appear to be biosynthetic in nature, as either one of the corresponding genes is able to complement a lack of DHOD activity in a *pyrD* deficient strain of *E. coli*. Furthermore, both genes must be

Figure 1



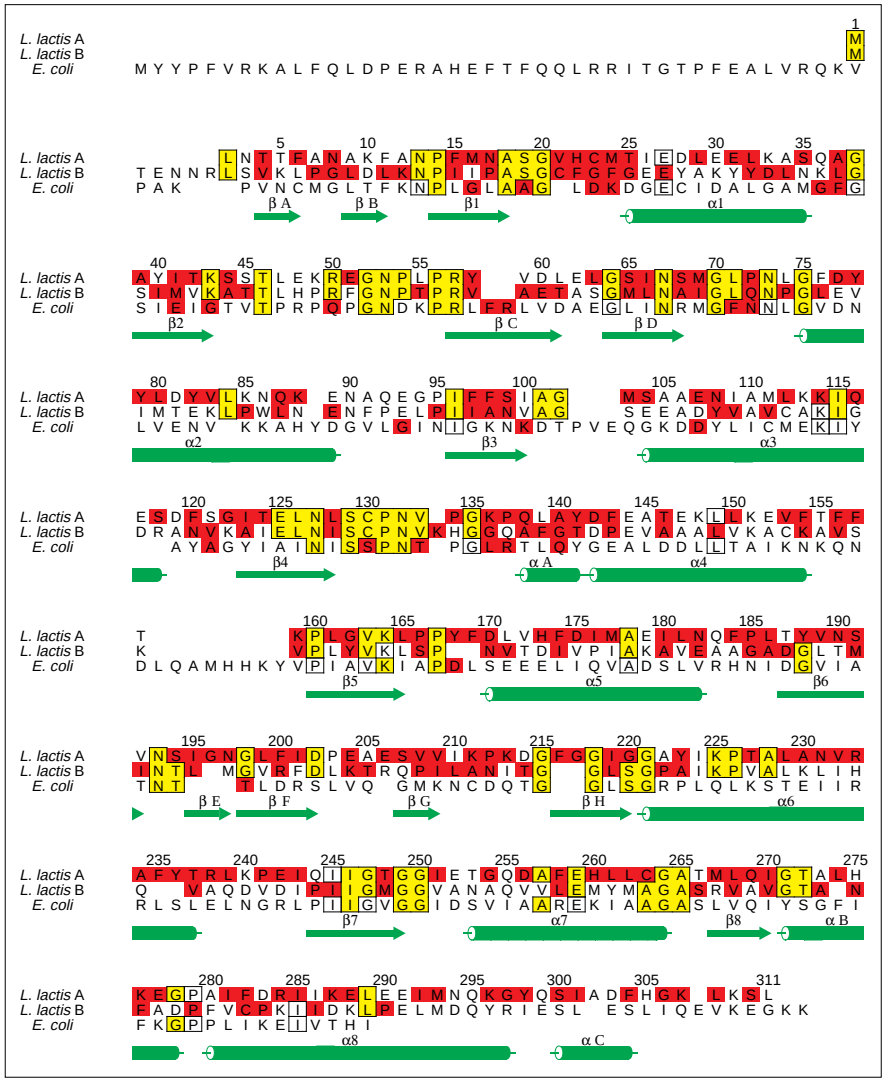
The reaction catalyzed by DHOD. The oxidation of the enzyme substrate (*S*)-dihydroorotate is accompanied by the reduction of the flavin prosthetic group. The flavin carbon atoms are labelled.

inactivated in *L. lactis* to impose a pyrimidine requirement on the organism [4]. As illustrated in Figure 2, the two enzymes show only 30% sequence identity though both proteins are polypeptides consisting of 311 amino acids. One of the enzymes, DHODA, is expected to be very similar to the enzyme from *S. cerevisiae*, the enzymes having a sequence identity of 71%. However, the other enzyme, DHODB, has only a 27% sequence identity with the *S. cerevisiae* enzyme and more closely resembles the enzymes from *Bacillus subtilis* (where the sequence identity is 66%) and several other Gram positive bacteria [5]. The currently known DHOD sequences can thus be divided into three families which can be represented by the two *L. lactis* enzymes and by the *E. coli* enzyme (Fig. 2). The known sequences of the mammalian DHODs are found in the latter family.

The two *L. lactis* DHOD enzymes both form dimers containing one molecule of FMN per subunit (~34 kDa), however, the DHODB enzyme is unstable and relatively inefficient when compared to DHODA [6]. Recently it has been found that the activity of the DHODB enzyme, both *in vivo* and *in vitro*, is dependent on the presence of an additional 262 residue protein encoded by a neighbouring open reading frame (*pyrK*) which is co-transcribed with *pyrDb* [4]. This protein contains a [2Fe–2S] iron-sulphur cluster and forms a complex tetramer with DHODB, consisting of two molecules of each protein. It has been proposed that this complex of two different proteins is the prototype of the DHODs for Gram-positive bacteria [6]. Both DHODA and DHODB and the *pyrK* encoded protein also have high sequence similarities (greater than 30% identity) with the much bigger (greater than 1000 residues) dihydropyrimidine dehydrogenases found in many organisms. These enzymes can reduce all natural pyrimidine bases, and contain both flavin and iron-sulphur clusters on a single protein chain [7].

At present there is great interest in investigating inhibitors of DHOD as potential therapeutic agents for treating diseases involving aberrant cell proliferation [8]. The enzyme has recently been implicated as a possible target of the novel immunosuppressive agent leflunomide [9,10]. This compound has shown promise in human clinical trials for rheumatoid arthritis, and in blocking rejection after allograft and xenograft transplantation in animals. The active agent of leflunomide has been shown to be a potent inhibitor of human DHOD. Brequinar sodium is another immunosuppressive agent [9] and is known to work by inhibiting mitochondrial DHOD, and consequently the production of UMP in *de novo* pyrimidine biosynthesis. The associated depletion of pyrimidine in the cells is then thought to result in the antiproliferative effects displayed.

Figure 2



Sequence alignment showing representatives of the three main DHOD families. The figure is based on a multiple alignment of the 17 currently known sequences. Pairwise alignments were used to determine a similarity tree relating the different sequences, which was then used to construct a multiple alignment using the program AMPS [34]. Sequence conservation between the families was investigated with AMAS [35]. All residues coloured red are conserved within a family and if the same residue is conserved in more than one family it is coloured yellow. Boxed uncoloured residues are conserved in the representative alignment but not across the whole family. The secondary structure elements of the *L. lactis* DHODA structure are indicated on the alignment; β strands are shown as arrows and α helices as cylinders. These were determined with the help of DSSP [36], although strand β2 technically has undefined secondary structure. Residue numbers on the alignment refer to the *L. lactis* DHODA sequence. The sequences represented by the alignment consist of the following: (1) *L. lactis* A and *S. cerevisiae*; (2) *L. lactis* B, *Enterococcus faecalis*, *B. subtilis* and *B. caldolyticus*; and (3) *E. coli*, *Mycobacterium leprae*, *Plasmodium falciparum*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Drosophila melanogaster*, rat, human, *Agrocybe aegerita*, *Haemophilus influenzae* and *Salmonella typhimurium*. (Figure produced using the program ALSCRIPT [37].)

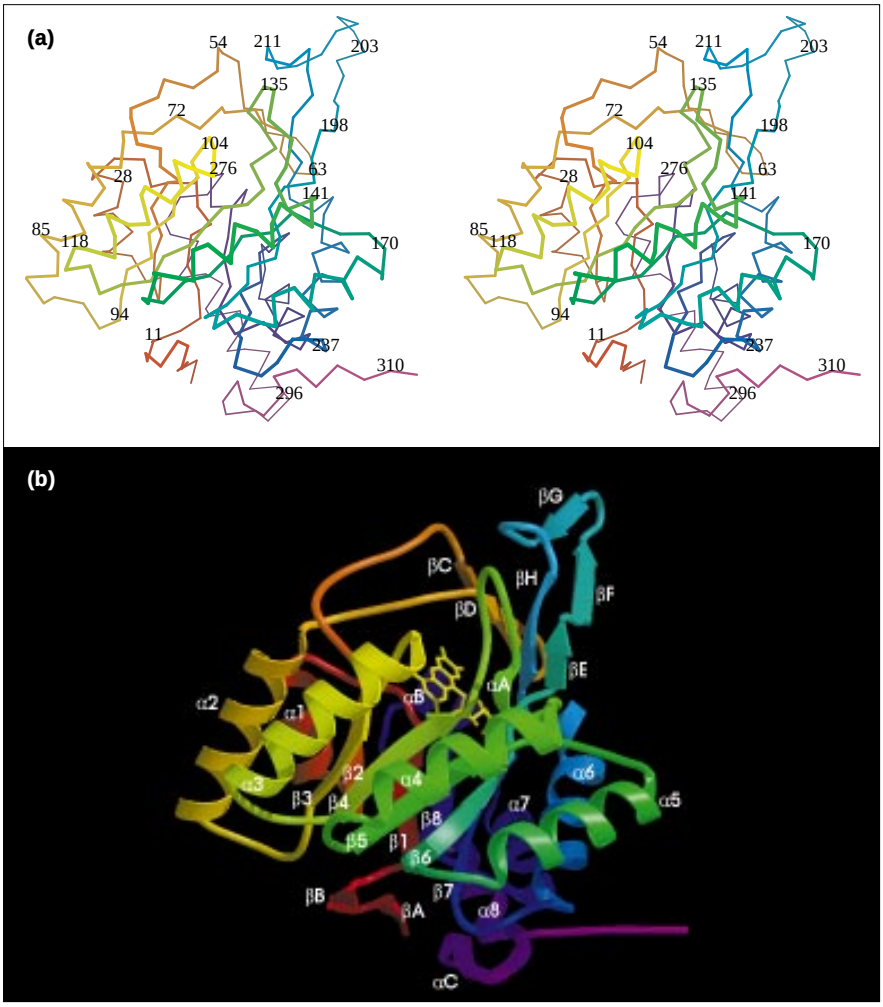
We are currently investigating the functional and structural properties of the three different families of DHODs using both biochemical and X-ray crystallographic methods. Here we present the first three-dimensional structure of a DHOD, the A form of the enzyme from *L. lactis*, which has been solved in its native form and refined to 2.0 Å resolution.

Results and discussion

The structure of the DHOD monomer

The *L. lactis* DHODA structure consists of a dimer of two identical subunits. The DHODA monomer is shown in Figure 3 and the assigned secondary structure is indicated in Figure 2. Like many other flavin containing proteins, the monomer has an α/β-barrel fold, consisting predominantly of a central core of eight parallel β strands (β1–β8) surrounded by a ring of eight α helices (α1–α8). Additional secondary structural elements (designated with letters) and loops are present as insertions towards the top of the barrel. The flavin molecule is located between the top of the barrel and the subdomain formed by these insertions. The barrel base is made up of two short antiparallel β strands (βA and βB) towards the N-terminal part of the protein. The N terminus itself is well ordered from the first residue in the sequence, packing against residues of helices α8 and αC. The mainchain hydrogen-bonding pattern of the central core is disturbed by the partial opening up of the barrel on either side of the second β strand (β2). The distance between β2 and its neighbours is much larger than the distance between the other strands, and its mainchain atoms make only one hydrogen bond with β1 and two hydrogen bonds with β3 that are shorter than 4 Å. Between β2 and α2 there is a long insert consisting of

Figure 3



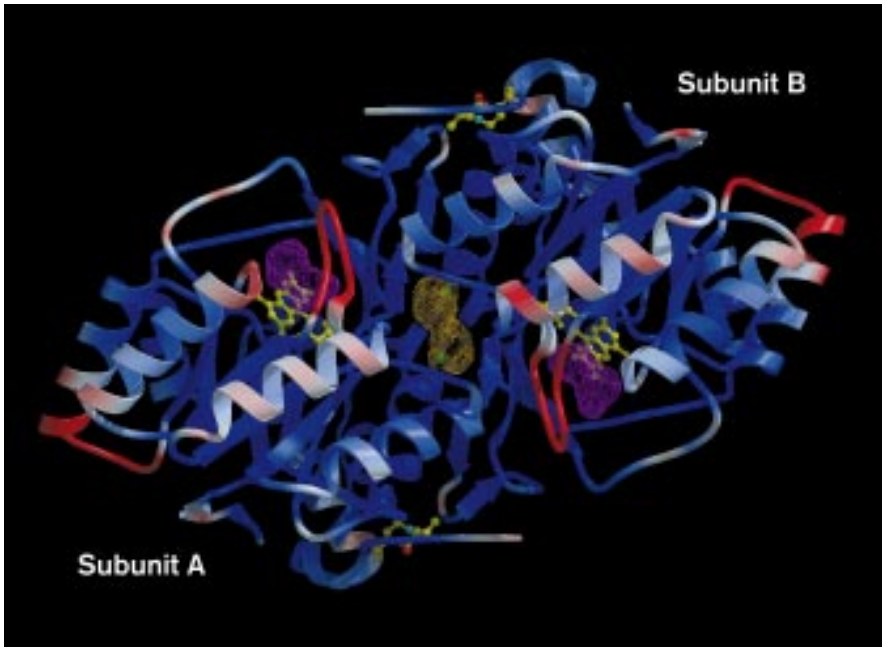
The monomer structure of DHODA from *L. lactis*. (**a**) Stereo diagram of the Cα trace of DHODA with some atoms labelled. (**b**) Secondary structure elements of the protein fold showing the location of the flavin group. Both representations are colour ramped from the N terminus (red) to the C terminus (magenta), and are shown in the same orientation. (Figure produced using the programs BOBSCRIPT [R Esnouf, unpublished program] modified from MOLSCRIPT [38], and RASTER3D [39].)

two highly conserved antiparallel loops which lead to strands βC and βD. Both loops and strands are approximately equidistant from each other for the length of the entire insert except at the point immediately leading into βC, where a short three-residue stretch (Pro54-Leu55-Pro56) bulges sharply out from the β2–βC loop. The peptide bond between the leucine and the conserved Pro56 is *cis*, and the following conserved residue (Arg57) is found along the line the mainchain would have taken if these three residues were absent. Following strand β4 there is another insert making up a highly conserved turn which leads into the short 3_{10} helix (αA) and then back into the barrel at α4. At the end of the sixth barrel strand there is a second *cis* peptide bond between Ser191 and Val192, neither of which are conserved. The third major insert into the barrel structure forms the four strands βE, βF, βG and βH; strands βH and βE lie alongside βC and βD making a four-stranded antiparallel β sheet. The two remaining strands βF and βG occupy the end of this insert. There is a sharp twist resembling a three-residue 3_{10} helix leading into the extremely glycine rich βH strand which rejoins the barrel at α6. After the last barrel strand there is an extra helix (αB) linked directly to α8 by a proline turn (Pro279). Helix α8 is the longest helix in the structure and one of the most variable regions amongst the sequences which are currently known. The last part of the monomer consists of the five residue 3_{10} helix (αC) and a long C-terminal tail. The C-terminal tail is very well defined even though it makes no contacts with its own subunit; it is held by interactions with residues from the other subunit within the dimer.

The orientation of the DHOD monomer as shown in Figure 3 can be used to illustrate its approximate dimensions. The diameter of the α/β barrel as measured between the Cα atoms of the last residues of helices α2 (Glu89) and α6 (Thr237) is 39 Å, and the distance between the Cα atoms of the N-terminal residue (Met1) and the first residue of βG (Ser207) is 49 Å.

Figure 4

The DHODA dimer structure viewed from above the twofold axis. The orientation of subunit A is the same as that of Figure 3. Secondary structure elements are coloured according to B factor: residues with B factors below 20 Å^2 are dark blue and residues with B factors above 40 Å^2 are bright red. The three cavities inside the dimer are shown as probe occupied volumes, as calculated by the program VOIDOO [40] using a probe radius of 1.2 Å. The two cavities above the flavin molecules are coloured magenta, while the intersubunit void is shown in orange. The two water molecules inside the central void are shown as green spheres. The intersubunit Glu206–Lys296 salt bridges are shown in ball-and-stick representation. (Figure produced using the programs BOBSCRIPT and RASTER3D.)



Within the monomer as a whole there are eight salt bridges where oppositely charged residues are within 4 Å of each other. Five of these are formed across different secondary structure elements and may play a role in stabilization of the structure: His22–Glu31 (2.8 Å, β1–α1 loop/α1); Asp175–Arg238 (2.7 Å, α5/α6–β7 loop); Lys225–Asp256 (2.8 Å, α6/α7); Lys225–Glu259 (2.8 Å, α6/α7); and Glu277–Arg284 (2.9 Å, αB/α8). The remaining three bridges occur on neighbouring turns within α helices: Glu30–Lys33 (3.5 Å, α1); Asp283–Lys287 (3.8 Å, α8); and Arg284–Glu288 (2.8 Å, α8). None of these salt bridges comprise residues that are conserved in all DHODs, although both residues in the Lys225–Glu259 bridge are conserved in the two *L. lactis* DHOD protein families. Of the 25 completely conserved residues in the monomer the majority are clustered around the flavin group, either situated in loops or towards the C-terminal ends of the barrel strands. The relatively low sequence conservation in the α/β barrel secondary structure elements implies that the conserved residues are not generally associated with stabilizing the protein fold, but are more concerned with protein function.

The structure of the DHOD dimer

The dimer of DHODA is shown in Figure 4. It contains two copies of the DHOD monomer, related by a non-crystallographic twofold axis parallel to the crystallographic a* axis, the intersubunit rotation angle being 180° within the estimated standard deviation. There are essentially no significant differences between the two subunits, and they can be overlaid using the non-crystallographic symmetry axis to give a root mean square (rms) deviation in Cα positions of only 0.10 Å. The largest dimension of the dimer is 85 Å, as measured between the Cα atoms of the two Glu89 residues, and the distance between the two C-terminal Cα atoms is 48 Å. The dimer has a volume of about 55 300 Å^3 and an accessible surface area of 23 100 Å^2, as determined using the program GRASP [11]. The accessible surface area of the monomer is 13 800 Å^2, indicating a buried surface area between monomers of 2 200 Å^2 (16% of the surface area of the monomer).

At the dimer interface there are 28 residues from each subunit within 3.6 Å of residues of the other monomer, none of which are completely conserved. Of these, 14 lie between the start of strand βE and the start of strand βH (residues 195–199, 201, 203–208, 214 and 216). These residues make contact with residues of the other subunit: 169 (β5–α5 loop), 226 and 230 (α6), 263 (α7), 296 (α8), 298 (α8–αC loop), and 309–311 (C terminus). In addition, helix α5 (residues 170 and 172) makes contact with residues of the other subunit: 137 (β4–αA loop) and 138 (αA). The only other residue involved in intersubunit contact is residue 223 (α6) which makes a contact with residue 226 of the other subunit. Only two residues are within 3.6 Å of their own counterparts in the opposite subunit (Phe169 and Tyr223); in the case of Phe169, the two phenyl rings stack against each other. The two C-terminal tails run parallel with the βG strands making main-chain hydrogen bonds, and result in a short third β strand alongside βF and βG in both subunits. There are also two intersubunit Glu206–Lys296 salt bridges (2.8 Å),

which are conserved in the DHODA sequence family. Thus overall, the dimer contacts are a mixture of both hydrophilic and hydrophobic interactions.

A rather large cavity is found at the centre of the dimer having a volume of approximately 109 Å^3. This cavity is empty except for two very well ordered water molecules: Wat1026 (subunit A, B factor = 21.4 Å^2) and Wat1015 (subunit B, B factor = 25.8 Å^2). These are involved in well defined hydrogen-bond interactions with the mainchain atoms Tyr223 O and Thr227 N and the sidechain atom Thr227 Oγ1, from both subunits. The remainder of the cavity is lined by atoms from the hydrophobic sidechains of Ile195, Ile219, Tyr223, Ile224, Lys225, Pro226 and Thr227 from both subunits. Though none of these residues are completely conserved, we note that the three isoleucine residues in DHODA always correspond either to isoleucine, leucine or valine in the other sequences.

There are two other cavities in the DHODA dimer which are also shown in Figure 4. These cavities are situated above the flavin isoalloxazine rings in both subunits and correspond to the expected substrate-binding sites, as described later. As can be seen from Figure 4, the two flavin-binding sites are well separated preventing any interaction between the two active sites in the dimer. Thus although dimer formation does not appear to be important for catalytic activity, it is certainly important for enzyme stability as seen by the presence of the hydrophobic cavity in the centre of the dimer.

The final model of the dimer contains 353 water molecules. Of these water molecules there are 131 in each subunit which have a corresponding water within 1 Å of the equivalent position in the other subunit. The rms distance between corresponding waters is 0.36 Å as determined with LSQMAN (G Kleywegt, unpublished program) showing that 74% of the waters are found at equivalent sites within the two subunits. The rms difference in B factors between equivalent waters is 6.0 Å^2.

Structural similarities with other proteins

The DHODA structure closely resembles other FMN binding flavoproteins, which is not surprising as most of those structures solved are also α/β barrels. However, as with many α/β-barrel structures, there is only a very low sequence similarity between DHODA and other proteins that adopt this fold. However, one region that is often at least partially conserved is the phosphate-binding site, located between β7 and α8, where the flavin phosphate group is bound. The presence of the phosphate-binding site motif in some DHOD sequences, together with secondary structure predictions revealing some alternating sheet and helical regions, has previously been used to propose an α/β-barrel fold for DHOD [12]. However, knowledge of the DHODA structure now allows a more direct comparison between the DHODs and other enzyme families.

A search for structurally similar proteins was carried out using the program DALI [13]. The highest structural similarity was found between DHODA and the FMN-binding domain of flavocytochrome b_2 (FCB) [14], where superposition of 234 Cα atoms from the two proteins gave an rms deviation of 3.0 Å, with a corresponding sequence identity between equivalenced residues of 13%. FCB is a two-domain protein which binds both an FMN molecule and a haem group, and catalyzes the oxidation of L-lactate to pyruvate with the subsequent transfer of electrons to cytochrome *c*. A high structural similarity was also found between DHODA and the FMN-binding domain of trimethylamine dehydrogenase (TMD) [15], where 233 Cα atoms could be superimposed to give an rms deviation of 3.2 Å (sequence identity between equivalenced residues 7%). TMD is much larger than DHODA, having three domains, and in addition to FMN it also contains a [4Fe–4S] iron-sulphur cluster. This enzyme catalyzes the oxidation of trimethylamine to dimethylamine and formaldehyde. DHODA is also rather similar in structure to old yellow enzyme (OYE) [16], where 217 Cα atoms could be superimposed to give an rms deviation of 3.0 Å and a corresponding sequence identity of 14%. The exact biological function of OYE is still unknown. The α/β barrels of FCB and TMD bear a very close resemblance to the DHODA barrel, with helices and strands matching very closely. At the barrel base especially, the loops between strands and helices are very similar in length and conformation. However at the top of the barrel, where the active-site loops of DHODA are located, there is very little similarity in the structures, this area being the interface between the different domains in FCB and TMD. The overall structure of the OYE barrel is again similar to that found in DHODA except that the length of the individual secondary structure elements varies markedly: in OYE many of the helices are longer than those in DHODA and others are much shorter. Although OYE contains a loop region at the top of the barrel, there is no resemblance between this region and the loop region found in DHODA. Thus, although there are structural similarities between these α/β barrels and the DHODA α/β barrel, there appears to be no correlation between the sequence identity of aligned residues and the corresponding structural similarity. Despite OYE exhibiting the greatest sequence identity between aligned residues with DHODA, the structure of this enzyme corresponds less well to DHODA than either FCB or TMD.

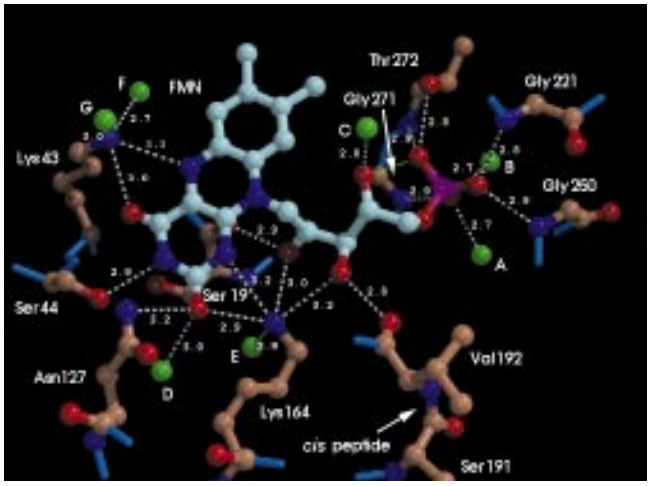
Among the 25 completely conserved residues in the DHOD sequences, there are seven which are also found in some of the three enzymes FCB, TMD and OYE. FCB from *S. cerevisiae* [14] has five residues in structurally equivalent positions conserved in DHOD (corresponding

to Pro14, Gly249, Gly250, Gly264 and Ala265 in DHODA), TMD from methylotrophic bacterium [15] has two (corresponding to Gly249 and Ala265 in DHODA), and OYE from *Saccharomyces carlsbergensis* [16] has three (corresponding to Gly75, Ile246 and Gly249 in DHODA). Of these residues, Ile246, Gly249, Gly250, Gly264 and Ala265 are all located in the phosphate-binding site region. The exact similarity in conformation and role of these residues clearly suggests that these proteins are at least distantly related to DHOD and supports the suggestion that enzymes with an α/β-barrel fold, having the conserved phosphate-binding site motif, have evolved by divergent evolution [12].

The flavin-binding site

The flavin molecule is very well buried within the protein, the only parts visible from the surface being a small region of the dimethylbenzene side of the isoalloxazine ring and the tip of the phosphate group. Figure 5 shows the flavin environment in subunit A with all its interactions to the protein atoms. The flavin-binding site is on top of the α/β barrel above the C-terminal ends of the barrel strands, with the *re*-face of the isoalloxazine ring resting on the top of the barrel. Consequently, most of the flavin interactions are made either with C-terminal residues of the β strands or with residues just beyond the C-terminal ends of the strands. Hydrogen bonds are formed between the FMN isoalloxazine ring and the residues Lys43 and Ser44 (end of β2), Asn127 (end of β4), Lys164 (end of β5) and with a water molecule. The dimethylbenzene hydrophobic part of the isoalloxazine ring system does not have any close contacts. The two methyl groups are loosely surrounded by carbon atoms from Val21 (β1–α1 loop), Tyr58 (start of βC), Asn 67 and Met69 (end of βD). Apart from the amide group of Asn67, which is 3.5 Å away, the distances to the surrounding mainly hydrophobic neighbours are all larger than 3.6 Å. The isoalloxazine ring is almost flat, the *re*-face being very slightly concave, as expected for an oxidized FMN molecule. The ribityl hydroxyl groups make hydrogen bonds to the carboxyl group of Ser19 (end of β1), the Lys164 ammonium group, the carboxyl group of Val192 (end of β6) and a water molecule. The *cis* peptide bond between Ser191 and Val192 facilitates the hydrogen bond to the second hydroxyl group O3′. As mentioned earlier, neither of the residues involved in this bond are conserved, however, as the interaction with the flavin involves a mainchain carboxyl group it may be that the nature of the sidechain of this residue is not that important. The protein interactions with the flavin phosphate group are dominated by mainchain nitrogen atoms, especially glycine nitrogen atoms. There are hydrogen bonds between the phosphate oxygen atoms and the mainchain NH groups of Gly221 (start of α6), Gly250 (end of β7), Gly271 and Thr272 (start of αB); the hydroxyl group of Thr272 also acts as a donor to the same oxygen atom that accepts a proton from its amide group. All the phosphate hydrogen bonds are short, being less than 3.0 Å.

Figure 5



Protein interactions with the flavin group. All residues making interactions with atoms of the flavin group within 3.6 Å are shown. Bonds drawn in light blue illustrate the continuation of the mainchain protein atoms on either side of all the residues shown; atoms are shown in standard colours. Hydrogen bonds between the protein and flavin are shown as dotted lines together with the interatomic distances in Å. Water molecules making hydrogen bonds with the flavin atoms are shown as green spheres: A (Wat1006, bound by Gly249 N, Ile251 N, Gln269 O); B (Wat1012, bound by Ile251 O, Ala273 N, Wat1091); C (Wat1022, bound by Ser194 Oγ, Wat1004); and D (Wat1008, bound by Ser99 Oγ, Glu125 Oε2). In the case of residues Lys164 and Lys43, three water molecules making hydrogen bonds with the lysine Nζ atoms are also shown: E (Wat1005, bound by Ser19 N, Glu125 Oε2, Gln269 Oε1); F (Wat1069, bound by Gly20 O); and G (Wat1246, bound by Leu71 N). All water molecules A–G are very well ordered, the mean B factor being 18.1 Å^2; water E has a B factor of only 12.1 Å^2. The *cis* peptide between Ser191 and Val192 is also shown. The dotted green line between two of the flavin oxygen atoms represents the extremely short intramolecular hydrogen bond of only 2.4 Å. (Figure produced using the programs BOBSCRIPT and RASTER3D.)

There are only four residues interacting with the flavin group that are conserved in all three families of DHODs, identified from the known sequences (Fig. 2): Asn67 and Asn127 (in contact with the isoalloxazine ring system) and Gly221 and Gly250 (in the phosphate-binding region). The remainder of the above mentioned residues directly surrounding the flavin group are conserved either among one or two of the three families: DHODA family only (Val21, Ser44, Tyr58, Met69 and Ser191); DHODA and DHODB families (Ser19, Lys43, Gly271 and Thr 272); and DHODA and *E. coli* families (Lys164).

The two nitrogen atoms in the isoalloxazine ring which change oxidation state during the reaction, N1 and N5 (Fig. 1), are hydrogen bonded to two lysine residues, Lys164 and Lys43, respectively. This suggests that these residues play an important role in the reaction mechanism, and consequently the hydrogen-bonding patterns of these residues deserve special attention. In the case of Lys164,

this residue acts as a donor in two bifurcated hydrogen bonds, one to ribityl oxygens O2′ and O3′ and the other to isoalloxazine ring atoms N1 and O2, and also makes a normal linear hydrogen bond to Wat1005 (water E in Fig. 5). Although Lys164 is not completely conserved (in *Bacillus caldolyticus* it is apparently replaced by an asparagine), the presence of a positively charged residue so close to the flavin N1 atom makes it likely that this nitrogen atom is unprotonated in the fully reduced flavin group. As a consequence the flavin group will be in the anionic hydroquinone form, as found in FCB [17]. In a similar manner to Lys164, Lys43 is involved in a bifurcated hydrogen bond (with isoalloxazine ring atoms O4 and N5) and makes two linear hydrogen bonds to two water molecules. This residue is only conserved among the *L. lactis* DHOD families. The equivalent residue in the *E. coli* enzyme family is a conserved glycine which would lead to a significant difference in the environment around flavin atom N5. This observation suggests significant differences in the catalytic mechanism of the two *L. lactis* enzymes and the *E. coli* enzyme.

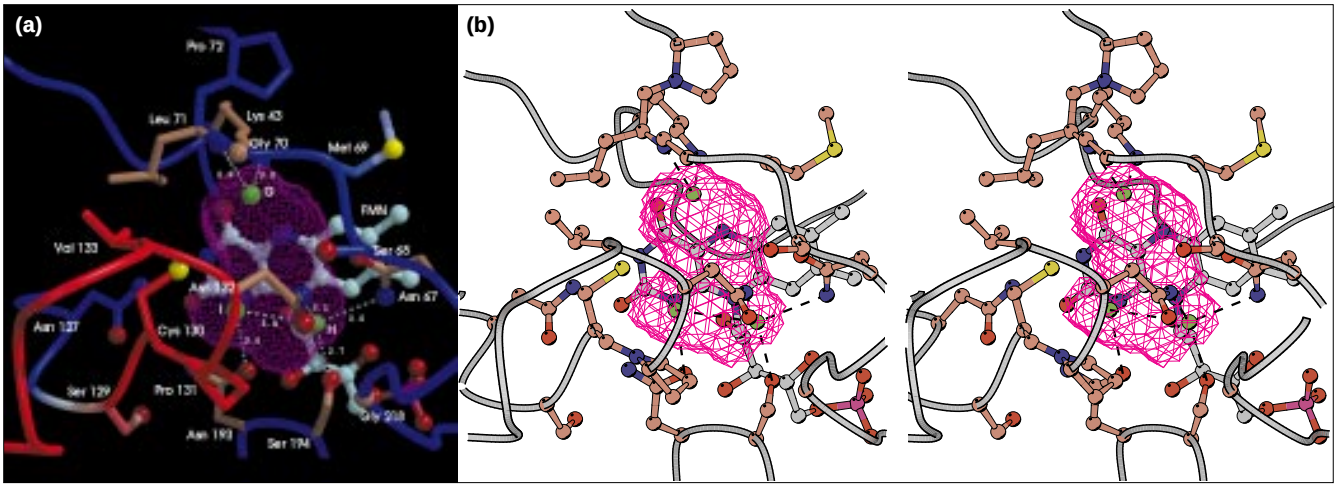
With this difference in mind it is interesting to compare the flavin environment in DHODA with those in the flavin enzymes that exert a similar catalytic function and showed resemblance with DHODA with respect to the folding of their α/β barrels. The FMN moiety adopts a conformation in DHODA that closely matches the conformations seen in FCB, TMD, OYE and also glycolate oxidase (GOX) [18]. Considering first the phosphate group, which is distant from the redox centre at the isoalloxazine ring system, we note that it is surrounded by residues with neutral sidechains in DHODA. This is in stark contrast to its environment in FCB, GOX, TMD and OYE, where the phosphate group is hydrogen bonded to a positively charged arginine sidechain. However, the short intramolecular hydrogen bond, between phosphate oxygen OF2 and ribityl hydroxyl oxygen O4′ in DHODA, that could assist in the stabilization of the phosphate negative charge is also observed in these related structures. More interesting are the similarities and differences in the protein interactions with the isoalloxazine ring system. Positively charged residues corresponding to Lys164 are found in all these related structures. In the case of FCB and GOX there is a lysine residue, and in TMD and OYE there is an arginine. In all these enzymes, the proximity of the lysine or arginine sidechain stabilizes a negatively charged N1 in the reduced flavin. The other nitrogen atom that changes its oxidation state during the enzyme catalyzed reaction is N5. The interaction seen between this atom and the positively charged Lys43 in DHODA is not found in any of the currently known FMN-binding protein structures determined to date. The hydrogen-bonding patterns of flavin atom N5 are much more variable amongst the other flavoproteins: N5 makes a hydrogen bond with a neutral mainchain NH group in TMD and OYE, a hydrogen bond with a water molecule in GOX, and a combination of these in FCB.

The substrate-binding site

From the structural studies of other FMN-binding enzymes the active site in a flavoprotein is known to be found close to the flavin isoalloxazine ring system. As mentioned previously, a rather large cavity containing three water molecules is located above both flavin ring systems in the DHODA structure. Both cavities are relatively large (102 $Å^3$ in both subunits) and are completely isolated from the dimer surface. The cavities lie beneath the β4–αA loops which are much more flexible than the majority of the rest of the protein atoms, as indicated by their thermal parameters (Fig. 4). It seems likely that access to or from these cavities would require some significant movement in the β4–αA surface loops. The cavity can be described as a flat circular disc centred above the connection between the first and second rings of the isoalloxazine system. The subunit A cavity and the residues surrounding it are shown in Figure 6. The surrounding residues come from strand β2 (Lys43), from strand βD and beyond (Asn67, Ser68, Met69, Gly70, Leu71 and Pro72), from strand β4 leading into the β4–αA loop (Asn127, Ser129, Cys130, Pro131, Asn132 and Val133), from the β6–βE loop (Asn193 and Ser194), and from strand βH (Gly218). Of these residues, all four asparagines, the two glycines, Ser129 and Pro131 are conserved in all known DHODs. Although Met69 is not conserved, it is virtually always a hydrophobic residue (methionine or isoleucine in the two *L. lactis* families, and generally tyrosine, methionine or cysteine in the *E. coli* family) as would be consistent with maintaining the hydrophobic environment of the dimethylbenzene part of the flavin group. The adjacent residue, Ser68, pointing in the opposite direction to Met69, is therefore unlikely to be concerned with the dimethylbenzene ring environment and perhaps for this reason is more variable (Fig. 2). The two hydrophobic residues, Leu71 and Val133, are always a leucine or phenylalanine in the case of the former, and either a valine or a threonine in the latter. Pro72 is also found in the neighbourhood, but although conserved within individual members of the three DHOD families, it is perhaps too far from the cavity to be of any significance to substrate binding. Ser194 is not conserved but is, nevertheless, either a serine or a threonine and is therefore likely to be involved in substrate binding.

The three water molecules within the cavity are found in identical positions in both subunits. Upon substrate binding they must leave the cavity. Nevertheless, their positions may indicate where the atoms of the substrate capable of forming hydrogen bonds could be found. Wat1047 (water H in Fig. 6) is bound by three well positioned protein residues and, in the fourth hydrogen-bond position, by a weaker hydrogen bond to Wat1108 (water I

Figure 6



The substrate-binding site. **(a)** The area around the subunit A substrate cavity. The cavity, shown as a probe accessible surface, is the same as that shown in Figure 4. The three water molecules inside the cavity (green spheres) are shown together with their hydrogen-bond interactions (dashed lines) and hydrogen-bond distances in Å. All residues with atoms bordering the cavity are shown coloured according to B factor, as in Figure 4, except for residues making direct hydrogen bonds with the cavity water molecules which are shown in standard atom colours. Of the three water molecules, water G (Wat1246, B factor = 26.1 Å^2) and water H (Wat1047, B factor = 30.0 Å^2) are the best defined. Water I (Wat1108, B factor = 51.6 Å^2) is less well defined but still clearly visible in the final $2F_o–F_c$ electron-density map. **(b)** Stereo diagram of the same region of the structure. For clarity, the cavity shown here is calculated with a coarser grid. (Figure produced using the programs BOBSCRIPT and RASTER3D.)

in Fig. 6). Wat1108 only makes one short hydrogen bond to the protein (to Asn193). These two water molecules are hydrogen bonded to three of the conserved asparagine residues and a serine, and take up the bulk of the space in the lower part of the cavity as viewed in Figure 6. However, an interesting feature concerning the lower part of the cavity was also observed for the partially conserved residue Cys130. Following the completion of refinement, the final $2F_o–F_c$ electron-density map and the final $F_o–F_c$ map still showed residual density beyond the cysteine SH group in both subunits, as evidenced in the original solvent flattened multiple isomorphous replacement (MIR) map. This electron density leads into the cavity next to the water molecule Wat1108. The residual density indicates that a partial oxidation of the sulphydryl group may have occurred during the crystallization period, indicating a high reactivity of this residue as the other cysteines in the structure do not show these features. An oxidized cysteine at this position would be able to make a hydrogen bond to the conserved Asn127, indicating that this latter residue could play a role in substrate binding.

The final water molecule bound in the cavity, Wat1246 (water G in Figs 5,6), is linked by two short hydrogen bonds to Lys43 Nζ and Leu71 N. Replacing this water molecule with the substrate must cause changes in the hydrogen-bonding pattern of Lys43. These changes could facilitate the subsequent reduction of the flavin group where N5 becomes protonated. As the substrate (*S*)-dihydroorotate is negatively charged the presence of any positively charged sidechains would be a good guide to predict the orientation of the substrate in the active site. Lys43 is the only positively charged residue close to the cavity, therefore it is likely to be responsible for the initial substrate interaction in DHODA. The location of water G, towards the top of the cavity as viewed in Figure 6, is significant in that the cavity bulges out away from the flavin ring system at this position (this is most easily seen in Figure 4 for the subunit B cavity). This observation suggests a possible location for the chiral centre of the substrate molecule, whereby the carboxylate group could point away from the isoalloxazine ring system filling the cavity at this point. Furthermore, if the chiral centre of the substrate occupied this position with the carboxylate group pointing away from the isoalloxazine ring system, then the methylene group which constitutes the only hydrophobic position in the substrate would be oriented towards the two hydrophobic sidechains of Leu71 and Val133, and the substrate hydrophilic groups would be positioned in the space occupied by the water molecules.

A structural alignment of the DHODA active site with the active sites of the two flavoproteins FCB and GOX, which have a similar catalytic function to DHOD, allows a direct comparison between the structures. The substrates for FCB and GOX (lactate and glycolate, respectively) are both α-hydroxy acids and are significantly smaller than (*S*)-dihydroorotate, the only known substrate for DHOD. (*S*)-dihydroorotate contains only one functional group and

has an overall hydrophilic ring system. The substrate orientation proposed above differs significantly from that found in FCB. In FCB the carboxylate group of the reaction product, pyruvate, has been shown to be hydrogen bonded to an arginine, which is also present in GOX leading to a prediction that this arginine has the same role in GOX substrate binding [17]. In DHOD, we find that the sidechain of the conserved Asn67, close to Lys43 and hydrogen bonded to water H, is similarly positioned relative to the flavin group, although the mainchain in DHOD differs considerably in relation to FCB and GOX in this region. A tyrosine residue, also found in both FCB and GOX, which makes a hydrogen bond to the pyruvate carboxylate group has no corresponding residue in DHODA due to major differences in the protein fold at this position. There are, however, some similar features between the substrate-binding sites of DHODA and those of FCB and GOX. The fully conserved residue Asn193 is found at the same position as the histidine residue hydrogen bonded to the pyruvate ketone carbonyl group in FCB. This carbonyl group is also hydrogen bonded to a tyrosine sidechain in FCB, which corresponds to Ala101 in the DHODA structure. The substitution of tyrosine with alanine could leave extra space for the larger dihydroorotate substrate.

Insights into the catalytic function of DHODs

Kinetic studies with bovine liver mitochondrial DHOD [19] have suggested possible mechanisms of catalysis. The mechanisms involve proton abstraction from the substrate methylene group leading to a hydride transfer from the chiral carbon atom in the substrate to the N5 of the flavin isoalloxazine ring (Fig. 1). As mentioned above, Lys43 in the two *L. lactis* sequence families is replaced by a glycine residue in the *E. coli* family. Given that this residue makes hydrogen bonds to the flavin N5 and O4 isoalloxazine ring atoms as well as water G in the DHODA enzyme, and probably hydrogen bonds to the substrate carboxylate group, differences in this residue must be of crucial importance for the detailed reaction mechanism in the different sequence families. Assuming that the presence of a positively charged sidechain is essential for the proper orientation of the substrate carboxylate group, the arginine residue that corresponds to Ser68 in DHODA may play the same role as Lys43 in the mitochondrial enzyme family (Fig. 2). Apart from the *Plasmodium falciparum* sequence, which differs in many other regions of the sequence, this residue is an arginine in all the known *E. coli* family sequences, which is closer to the situation seen in FCB and GOX.

If the enzyme mechanism proceeds by a proton abstraction from the substrate methylene group and the substrate orientation in the active site is as we predict, then we would expect to find a residue capable of accepting a proton in the immediate vicinity. The only obvious candidate for this is Cys130. This residue is conserved among the DHODA and DHODB families, but is replaced by a serine in the *E. coli* family. There is clearly little similarity with the reaction of FCB and GOX which utilize a histidine residue as the base. There are no negatively charged residues nearby that could lower the pK_a value of Cys130, but binding of the substrate could have the same effect, making Cys130 assume its basic form in the pH range 7.5–9.0 (the optimum pH range of the enzyme). Interestingly, a short sequence around Cys130 in DHODA (Glu-Leu-Asn-Leu-Ser-Cys-Pro) is conserved in the dihydropyrimidine dehydrogenases [7]. This sequence was identified as being part of the uracil-binding site in the dihydropyrimidine dehydrogenases, because the irreversible inactivation of the bovine enzyme with 5-iodouracil and NADPH led to the formation of an (*S*)-hexahydro-2,4-dioxo-5-pyrimidinyl derivative of the cysteine [20]. Furthermore, it was suggested that the thiol group could function as a general acid required to protonate uracil upon reduction by flavin [21].

The sequence conservation between the three enzyme families shows that the DHODA and DHODB families are most closely related. Nevertheless, because the *L. lactis* DHODB enzyme requires the [2Fe–2S] iron-sulphur cluster containing *pyrK* encoded protein to be present for optimum activity the two enzyme families clearly have some differences in their reaction mechanism. The similarity in structures between the DHODA enzyme and TMD is significant in relation to the complex formed with the *pyrK* encoded protein. A search of protein sequences with the program BLAST [22] revealed no significant sequence similarity between the *pyrK* encoded protein and TMD. However, a three-dimensional structure prediction based on the threading technique [23], carried out with the program TOPITS [24], gave reasonably high similarity scores with, firstly, phthalate dioxygenase reductase (PDR), and secondly the C-terminal half of TMD. Threading of the *pyrK* protein sequence onto the PDR and TMD structures leads to a 19% sequence identity between aligned residues in both cases. The PDR enzyme contains a [2Fe–2S] cluster, as found in the *pyrK* encoded protein, in contrast to the TMD enzyme which has a [4Fe–4S] cluster. Although these two enzymes differ substantially as a whole, PDR contains an NAD-binding domain which is rather similar to one of the TMD domains. Given that the *pyrK* encoded protein contains a [2Fe–2S] cluster and can bind NAD [6] it thus seems likely that the DHODB–PyrK protein complex would consist of a DHODA-like domain and a PDR-like domain, with the domain interface at the top of the α/β barrel amongst the loop regions. Furthermore, if the similarities with TMD are significant then the TMD structure may suggest the orientation between the α/β barrel and the other iron-sulphur cluster containing domain. The high sequence similarity between DHODs and dihydropyrimidine dehydrogenases (which also have similarity to the

pyrK encoded protein, as described earlier) also suggests that two of the domains in dihydropyrimidine dehydrogenases could be a DHODA-like domain and possibly a PDR-like domain, again resembling in part the DHODB–PyrK protein complex.

The crystal structure of DHODA thus provides a good basis for investigating the mechanism of this enzyme by mutant studies and for studying the differences in catalytic function in the different DHOD families. The structure may also be useful for modelling the DHODB–PyrK protein complex and two of the domains in the much larger dihydropyrimidine dehydrogenases.

Biological implications

Dihydroorotate dehydrogenase (DHOD) is a flavoprotein oxidoreductase, which catalyzes the oxidation of (*S*)-dihydroorotate to orotate, the fourth step in the *de novo* biosynthesis of pyrimidine nucleotides. Inhibition of the mitochondrial human enzyme leads to dramatic reductions in cellular levels of pyrimidines. This inhibition is currently under investigation as a means of treating diseases involving aberrant cell proliferation. DHOD is closely related to the much larger dihydropyrimidine dehydrogenases, which can reduce all natural pyrimidine bases and the synthetic analogues used in chemotherapy. *Lactococcus lactis* contains two different functional DHOD proteins, one of which shows optimal activity only in the presence of a second, co-transcribed iron-sulphur cluster containing protein, with which it forms a complex tetramer. There are three families of DHODs based on the two *L. lactis* families and the family of membrane bound proteins represented by the *E. coli* enzyme, which includes the enzymes of mitochondrial origin. Investigation of the similarities and differences between the two *L. lactis* enzymes and the mode of association in the complex should allow some understanding of the reasons for the two different DHOD proteins. The differences in these enzymes also make them particularly interesting for studying flavoprotein redox reactions involving iron-sulphur clusters, and may suggest explanations for having a complex rather than a single chain polypeptide, such as that found in dihydropyrimidine dehydrogenases.

The crystal structure of DHODA is the first structure reported for this class of enzymes. It explains the function of many of the conserved residues, and has allowed potential active-site residues to be identified. The flavin-binding site shows many features which are common to other flavin mononucleotide containing proteins, but also reveals two new features: a phosphate-binding site without charged residues and a sidechain residue (Lys43) which interacts with the central ring of the flavin group. The substrate-binding site has been iden-tified as a water filled cavity completely buried underneath a highly conserved and apparently flexible loop. The cavity is surrounded mostly by conserved residues, although two of the residues in the active site (Lys43 and Cys130) which are conserved in the two *L. lactis* DHOD families are replaced by a glycine and a serine, respectively, in the mitochondrial enzyme family. This observation suggests significantly different mechanisms for these enzymes, and Cys130 is proposed as the catalytic base in the DHODA family. Thus this crystal structure provides a good basis for investigating the reaction mechanism of DHOD in detail and for understanding the differences between the three DHOD sequence families.

Materials and methods

Crystallization

Bright yellow crystals of *L. lactis* DHODA were obtained using the hanging drop vapour diffusion technique from solutions containing 30% PEG (4000, 6000 or 8000), 0.2 M sodium acetate and 0.1 M Tris-HCl, pH 8.5, as described previously [5]. The crystals are monoclinic in space group $P2_1$ with unit cell dimensions a = 54.2 Å, b = 109.2 Å, c = 67.2 Å and β = 104.5°. There are two molecules in the asymmetric unit, giving an estimated crystal solvent content of 56%. The two molecules are related by a non-crystallographic twofold axis parallel to the crystallographic a* axis. Heavy-atom derivatives were prepared by adding various amounts of a solution of 0.5 M $KAu(CN)_2$ in water directly to drops containing crystals in their original mother liquor.

Data collection and processing

All data sets were collected in-house on an R-axis II imaging plate system with a Rigaku RU200 rotating anode CuK_α X-ray source (λ = 1.54 Å) operating at 50 kV and 180 mA, using a graphite monochromator and a 0.5 mm collimator. The native data set was collected from a single crystal kept at 15°C of approximate size 0.5 × 0.5 × 0.5 mm^3. A total of 72 diffraction images were collected, each covering an oscillation range of 2.5° and an exposure time of 30 min, giving an overall rotation range of 180°. These images were processed using the programs DENZO and SCALEPACK [25]. Structure factors were derived from the reflection intensities using the program TRUNCATE [26]. The merged native data set has an R_{merge} of 5.2% and is 99.3% complete to 2.0 Å.

The gold derivative data sets were collected from four crystals of varying sizes. Different heavy-atom concentrations and soaking times were used to prepare the crystals, as shown in Table 1. The data sets were processed using DENZO and the CCP4 program suite [26]. Data collection statistics for the native and derivative data sets are given in Table 1.

Structure determination

The structure was solved by the MIR method, using a single heavy-atom compound to achieve observable differences in the structure-factor amplitudes. The initial data set obtained from a heavy-atom soak with $KAu(CN)_2$ showed only a very low mean fractional isomorphous difference between the native and derivative data sets. However, a significant peak corresponding to a gold site could still be seen in the 2.8 Å difference Patterson, and a second, lower occupancy site could also be tentatively identified. Attempts to obtain higher occupancy gold binding resulted in three further data sets. All four data sets showed the same two gold sites, A and B, which were consistent with the non-crystallographic twofold axis. In the 2.5 Å difference Patterson, for the best data set (Gold 4), the two self vector peaks for sites A and B in the v = ½ Harker section had heights of 21σ and 19σ, respectively, and even in the corresponding 2.5 Å anomalous Patterson they were visible at heights of 10σ and 11σ, respectively. In the second data set (Gold 2) two further pairs of gold sites, also related by the same

Table 1

Data collection and phasing statistics.

	Native	Gold 1	Gold 2	Gold 3	Gold 4
Crystal size (mm^3)	0.5×0.5×0.5	1.0×0.8×0.2	0.5×0.4×0.2	0.8×0.8×0.2	0.4×0.4×0.4
Protein concentration ($mg\,ml^{-1}$)*	18	30	40	18	18
PEG component†	PEG 6000	PEG 4000	PEG 4000	PEG 6000	PEG 8000
$KAu(CN)_2$ soaking conditions		2.5 mM, 1 day	8 mM, 2 days	50 mM, 1 day	50 mM, 2 days
Total data collected (°)	180	180	120	180	160
Oscillation range per frame (°)	2.5	3	3	3	2
Exposure time per frame (min)	30	30	60	30	30
Resolution (Å)	2.0	2.8	2.8	2.8	2.5
Observations	179 624	67 880	44 519	68 497	80 280
Unique reflections	50 757	18 583	16 641	18 804	25 489
Completeness (%) (all data / $I > 3\sigma I$)	99.3 / 80.3	99 / 92	89 / 56	99 / 89	96 / 85
R_{merge} (%)††	5.2	5.4	13.6	6.9	6.1
R_{iso} (%)§		7.8	18.4	13.2	19.3
Highest resolution shell (Å)	2.03–2.00	2.85–2.80	2.85–2.80	2.85–2.80	2.54–2.50
Completeness (%) (all data / $I > 3\sigma I$)	90.8 / 45.4	90 / 75	81 / 25	88 / 66	88 / 61
R_{merge} (%)	23.4	9.3	38.8	11.4	11.2
Heavy-atom sites		2	6	2	2
Phasing power (acentric/centric)‡		0.75 / 0.48	1.48 / 0.95	2.80 / 1.90	2.62 / 1.86
R_{cullis} (acentric/centric)#		0.91 / 0.96	0.75 / 0.79	0.51 / 0.54	0.54 / 0.54

*The protein concentration in the mother liquor. †The PEG component used in the mother liquor. ††$R_{merge} = \Sigma|Ij - <Ij>| / \Sigma <Ij>$, where Ij is the intensity of an observation of reflection j and $<Ij>$ is the average intensity for reflection j. §$R_{iso} = \Sigma||F_{PH}| - |F_P|| / \Sigma|F_P|$, where F_{PH} is the structure-factor amplitude of the derivative crystal and F_P is that of the native crystal. ‡Phasing power = root mean square ($|F_H| / E$), where F_H is the calculated structure-factor amplitude due to scattering by the heavy atoms and E is the residual lack of closure error. #R_{cullis} = lack of closure/isomorphous difference.

non-crystallographic twofold axis, were easily identified from a difference Fourier calculated using the first pair of sites for phasing. There was no sign of these extra sites in difference Fouriers calculated for the other data sets. It is unclear why these sites were not occupied in any of the other crystals, although it seems likely that the different soaking conditions or crystal thicknesses must have had some effect.

The derivative data sets were scaled to the native 2 Å data set using the program SCALEIT [26]. The heavy-atom positions, occupancies and temperature factors were refined for all the heavy-atom sites of all the derivatives simultaneously using the program MLPHARE [26]. All the data from 25 Å to the maximum resolution of each data set was used, anomalous pairs being included for all the derivatives. The derivative phasing powers and cullis R values are shown in Table 1. The final heavy-atom positions, occupancies and temperature factors for all the gold sites are given in Table 2. After phasing, the final overall mean figure of merit for all data from 25–2.5 Å was 0.56.

The MIR phases were improved with the program DM [26] using 20 cycles of solvent flattening (55% solvent content), histogram matching and Sayre's equation with phase extension from 2.5–2.0 Å. The resulting 2 Å electron-density map was easily interpretable, with α-helix directions, sidechains, carbonyl oxygens and even many water molecules easily identifiable. The map was skeletonized with BONES [27] and the protein model built using O [28]. The tyrosine residues were especially unmistakable and were used to assign the exact residue sequence numbers to the protein density. Figure 7a shows the experimental MIR electron-density map for helix α2, where the Tyr-Tyr-X-X-Tyr motif is clearly visible. Although in principle twofold averaging could have been used to improve the phases further it was not needed, and the protein could be traced from the first to the last residue with just two short loops (residues 88–94 and 133–137) missing in the initial model. Only one subunit was built into the density, with the non-crystallographic twofold axis being used to determine coordinates for the second subunit. Structure factors for the unrefined protein atoms were calculated and used to generate an F_o–F_c map, with peaks above 3σ being used to check for water molecules. A total of 72 well defined peaks were accepted which were all in very favourable hydrogen-bonding locations and were present in both the F_o–F_c map (above a 3σ level) and in the experimental map (above 1σ). Figure 7b shows the experimental MIR electron-density map in the area around the flavin group, where many well bound water molecules are clearly visible.

Refinement

Refinement was carried out using X-PLOR [29] at 2.0 Å resolution with all measured reflections being included without a σ cut. Free R factor test set reflections were selected with DATAMAN (G Kleywegt, unpublished

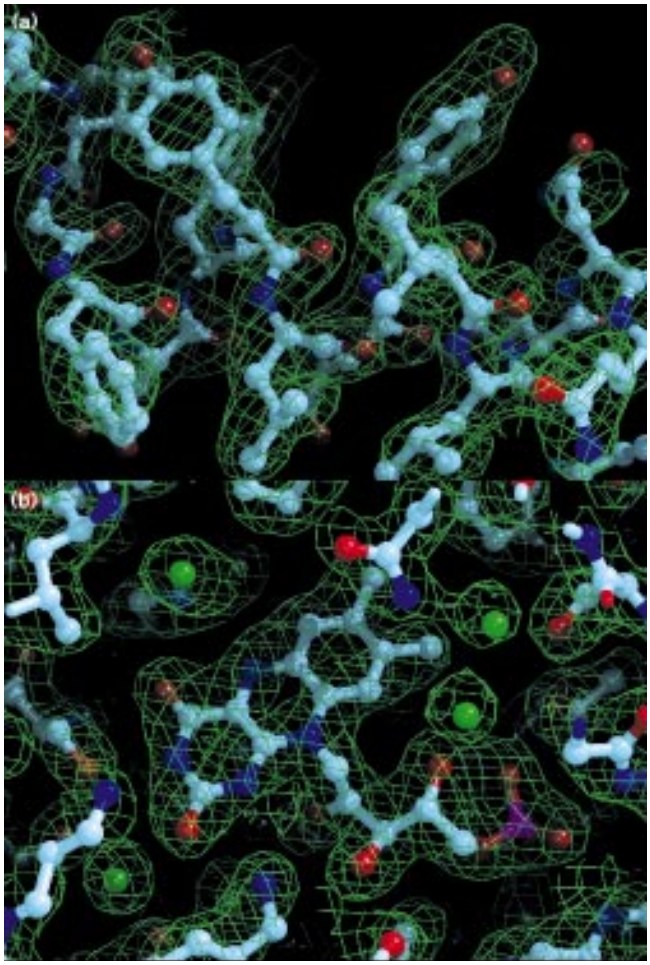
Table 2

Refined heavy-atom parameters.

Derivative	Site*	Fractional coordinates	Occupancy	B factor ($Å^2$)
Gold 1	A	0.928, 0.999, 0.784	0.22	23.6
	B	0.927, 0.586, 0.688	0.17	18.8
Gold 2	A	0.929, 0.999, 0.787	0.58	12.6
	B	0.927, 0.587, 0.684	0.53	7.8
	C	0.545, 0.931, 0.715	0.22	21.2
	D	0.546, 0.657, 0.603	0.20	35.1
	E	0.808, 0.938, 0.784	0.17	21.2
	F	0.808, 0.647, 0.639	0.19	23.9
Gold 3	A	0.929, 0.000, 0.786	0.66	20.1
	B	0.927, 0.587, 0.685	0.61	17.0
Gold 4	A	0.929, 0.000, 0.788	0.95	20.5
	B	0.927, 0.588, 0.683	0.89	17.4

*Gold atoms in sites A and B are bound to Cys23 residues; sites E and F correspond to binding to Cys130 residues. There are no obvious candidate residues for sites C and D.

Figure 7



The experimental 2.0 Å MIR solvent flattened electron-density map. (a) Helix α2 where the protein sequence was first matched to the electron density. The tyrosine residues are clearly visible and the well defined carbonyl oxygens allowed easy determination of the helix direction. (b) The area around the flavin group. This is one of the most well defined regions of the electron-density map, where many ordered water molecules could easily be discerned. Both maps are contoured at a 1σ level, where σ is the root mean square deviation of the electron density in the asymmetric unit. The final protein model is superimposed on the maps in both cases; atoms are shown in standard colours. (Figure produced using the programs BOBSCRIPT and RASTER3D.)

program) using the thin shells method [30] with 4% of the total reflections being included. At the start of refinement the R factor for reflections from 8–2 Å was 38.3% (free R = 38.4%). The model was first refined with very high non-crystallographic symmetry restraints using positional refinement, simulated annealing and individual B factor refinement. After this the R factor had dropped to 23.2% (free R = 25.4%). Following the inclusion of a bulk solvent correction and further positional and B factor refinements, the R factor had reached 21.1% (free R = 22.8%) for all reflections from 25–2 Å. At this point the protein model for subunit A was adjusted and the two missing loops could now be inserted into the electron density. Again, the coordinates for subunit B were generated from the non-crystallographic symmetry matrix. All of the MIR map assigned waters were carefully checked to ensure there was still good electron density present (which was the case for all 72 waters) and some new waters were added. Two further cycles of refinement and rebuilding were carried out, during which all non-crystallographic symmetry restraints were removed. Finally, the hydrogen-bonding patterns of the glutamine, asparagine and histidine sidechains were examined with the help of the program WHATIF [31], and adjusted if necessary. The refinement was finished off with a positional refinement and a last B factor refinement. The R factor for all data up to 2 Å is 16.8% (free R = 21.2%); Table 3 gives a summary of the final refinement statistics. The final model contains all 311 residues of the protein in each subunit together with their corresponding flavin groups, and a total of 353 water molecules. Over 90% of the residues fall into the most favoured regions of the Ramachandran plot as defined by PROCHECK [32], and there are no residues in disallowed regions. The mean coordinate error based on the SIGMAA method [33] is about 0.2 Å.

Table 3

Refinement summary.

Resolution range (Å)*	25–2.0
R factor (%)†	16.8
Free R factor (%)‡	21.2
Non-hydrogen protein atoms	4818
Flavin atoms	62
Waters	353
Solvent content (%)	57.6
Intersubunit rotation angle (°)	179.94
Rms deviations from ideality	
bond lengths (Å)	0.010
bond angles (°)	1.591
Average B factors (Å^2)	
mainchain	
(subunit A/subunit B)	24.4/24.3
sidechain	
(subunit A/subunit B)	27.6/27.4
flavin atoms	16.0
waters	39.3
Rms deviations in B factors (Å^2)§	
mainchain atoms	1.9
all atoms	2.3

*The resolution range is for 50 732 reflections; †the R factor was calculated using 48 672 reflections; ‡the free R factor was calculated using 2 060 reflections. §The rms deviations refer to bonded atoms.

Accession numbers

The atomic coordinates and structure-factor amplitudes of DHODA have been deposited with the Brookhaven Protein Data Bank; accession numbers 1DOR and R1DORSF.

Acknowledgements

We would like to thank Anders Kadziola, Thomas N Petersen and Olof Björnberg for many helpful discussions. We are grateful to the Danish National Research Foundation for the funding of this research.

References

1. Larsen, J.N. & Jensen, K.F. (1985). Nucleotide sequence of the *pyrD* gene of *Escherichia coli* and characterization of the flavoprotein dihydroorotate dehydrogenase. *Eur. J. Biochem.* **151**, 59–65.
2. Andersen, P.S., Jansen, P.J.G. & Hammer, K. (1994). Two different dihydroorotate dehydrogenases in *Lactococcus lactis*. *J. Bacteriol.* **176**, 3975–3982.
3. Roy, A. (1992). Nucleotide sequence of the URA1 gene of *Saccharomyces cerevisiae*. *Gene* **118**, 149–150.
4. Andersen, P.S., Martinussen, J. & Hammer, K. (1996). Sequence analysis and identification of the *pyrKDbF* operon from *Lactococcus lactis* including a novel gene, *pyrK*, involved in pyrimidine biosynthesis. *J. Bacteriol.* **178**, 5005–5012.
5. Nielsen, F.S., Rowland, P., Larsen, S. & Jensen, K.F. (1996). Purification and characterization of dihydroorotate dehydrogenase A from *Lactococcus lactis*, crystallization and preliminary X-ray diffraction studies of the enzyme. *Protein Sci.* **5**, 852–856.

6. Nielsen, F.S., Andersen, P.S. & Jensen, K.F. (1996). The B form of dihydroorotate dehydrogenase from *Lactococcus lactis* consists of two different subunits, encoded by the *pyrDb* and *pyrK* genes, and contains FMN, FAD, and [FeS] redox centres. *J. Biol. Chem.* 271, 29359–29365.
7. Yokota, H., *et al*., & Gonzalez, F.J. (1994). cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria. *J. Biol. Chem.* 269, 23192–23196.
8. Copeland, R.A., Davis, J.P., Dowling, R.L., Lombardo, D., Murphy, K.B. & Patterson, T.A. (1995). Recombinant human dihydroorotate dehydrogenase: expression, purification, and characterization of a catalytically functional truncated enzyme. *Arch. Biochem. Biophys.* 323, 79–86.
9. Davis, J.P., Cain, G.A., Pitts, W.J., Magolda, R.L. & Copeland, R.A. (1996). The immunosuppressive metabolite of leflunomide is a potent inhibitor of human dihydroorotate dehydrogenase. *Biochemistry* 35, 1270–1273.
10. Williamson, R.A., *et al*., & Ruuth, E. (1995). Dihydroorotate dehydrogenase is a high affinity binding protein for A77 1726 and mediator of a range of biological effects of the immunomodulatory compound. *J. Biol. Chem.* 270, 22467–22472.
11. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281–296.
12. Bork, P., Gellerich, J., Groth, H., Hooft, R. & Martin, F. (1995). Divergent evolution of a β/α-barrel subclass: detection of numerous phosphate-binding sites by motif search. *Protein Sci.* 4, 268–274.
13. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138.
14. Xia, Z.-X. & Mathews, F.S. (1990). Molecular structure of flavocytochrome b_2 at 2.4 Å resolution. *J. Mol. Biol.* 212, 837–863.
15. Lim, L.W., Shamala, N., Mathews, F.S., Steenkamp, D.J., Hamlin, R. & Xuong, N.H. (1986). Three-dimensional structure of the iron-sulfur flavoprotein trimethylamine dehydrogenase at 2.4 Å resolution. *J. Biol. Chem.* 261, 15140–15146.
16. Fox, K.M. & Karplus, P.A. (1994). Old yellow enzyme at 2 Å resolution: overall structure, ligand binding, and comparison with related flavoproteins. *Structure* 2, 1089–1105.
17. Lindqvist, Y., Brändén, C.-I., Mathews, F.S. & Lederer, F. (1991). Spinach glycolate oxidase and yeast flavocytochrome b_2 are structurally homologous and evolutionarily related enzymes with distinctly different function and flavin mononucleotide binding. *J. Biol. Chem.* 266, 3198–3207.
18. Lindqvist, Y. (1989). Refined structure of spinach glycolate oxidase at 2 Å resolution. *J. Mol. Biol.* 209, 151–166.
19. Hines, V. & Johnston, M. (1989). Mechanistic studies on the bovine liver mitochondrial dihydroorotate dehydrogenase using kinetic deuterium isotope effects. *Biochemistry* 28, 1227–1234.
20. Porter, D.J.T., Chestnut, W.G., Taylor, L.C.E., Merrill, B.M. & Spector, T. (1991). Inactivation of dihydropyrimidine dehydrogenase by 5-iodouracil. *J. Biol. Chem.* 266, 19988–19994.
21. Podschun, B., Jahnke, K., Schnackerz, K.D. & Cook, P.F. (1993). Acid base catalytic mechanism of the dihydropyrimidine dehydrogenase from pH studies. *J. Biol. Chem.* 268, 3407–3413.
22. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
23. Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* 358, 86–89.
24. Rost, B. (1995). TOPITS: threading one-dimensional predictions into three-dimensional structures. In *The Third International Conference on Intelligent Systems for Molecular Biology (ISMB)*. (Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T. & Wodak, S., eds), pp. 314–321, AAAI Press, Menlo Park, USA.
25. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing. Proceedings of the CCP4 Study Weekend*. (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
26. Collaborative Computational Project No.4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 50, 760–763.
27. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* 47, 110–119.
28. Jones, T.A. & Kjeldgaard, M. (1993). O - The manual. Version 5.9. Uppsala University, Sweden.
29. Brünger, A.T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, USA.
30. Kleywegt, G.J. & Jones, T.A. (1995). Where freedom is given, liberties are taken. *Structure* 3, 535–540.
31. Vriend, G. (1990). WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.* 8, 52–56.
32. Morris, A.L., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12, 345–364.
33. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* 42, 140–149.
34. Barton, G.J. (1990). Protein multiple sequence alignment and flexible pattern matching. *Methods Enzymol.* 183, 403–428.
35. Livingstone, C.D. & Barton, G.J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci.* 9, 745–756.
36. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
37. Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* 6, 37–40.
38. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946–950.
39. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Cryst. D* 50, 869–873.
40. Kleywegt, G.J. & Jones, T.A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst. D* 50, 178–185.